arXiv:1503.08844v1 [math.ST] 30 Mar 2015

New Fréchet features for random distributions and associated sensitivity indices

Jean-Claude Fort[a] and Thierry Klein [b**]

April 1, 2015

Abstract

In this article we define new Fréchet features for random cumulative distribution functions using contrast. These contrasts allow to construct Wasserstein costs and our new features minimize the average costs as the Fréchet mean minimizes the mean square Wasserstein$_2$ distance. An example of new features is the median, and more generally the quantiles. From these definitions, we are able to define sensitivity indices when the random distribution is the output of a stochastic code. Associated to the Fréchet mean we extend the Sobol indices, and in general the indices associated to a contrast that we previously proposed.

keywords contrats, Wassertein costs, sensitivity index

Introduction

Nowadays the output of many computer codes is not only a real multidimensional variable but frequently a function computed on so many points that it can be considered as a functional output. In particular this function may be the density or the cumulative distribution function ($c.d.f$) of a real random variable (phenomenon). In this article we focused on the case of a $c.d.f$ output. To analyze such outputs one needs to choose a distance to compare various $c.d.f.$. Among the large possibilities offered by the literature we have chosen the Wasserstein distances (for more details on wasserstein distances we refer to [**?**]). Actually for one dimensional probability distributions the Wasserstein$_p$ distance simply is the L^p distance of simulated random variables from a universal (uniform on $[0,1]$) simulator U: $W_p^p(F,G) = \int_0^1 |F^-(u) - G^-(u)|^p du = \mathbb{E}|F^-(U) - G^-(U)|^p$, where F^- is the generalized inverse of F. This means that using Wasserstein distances is to compare various $c.d.f$ from various codes on a same simulation space, which seems very natural in many situations. The most relevant cases seem to be $p = 2$ and $p = 1$, and in this paper we will work with.

*[a]MAP5 Université Paris Descartes, SPC, 45 rue des Saints Pères, 75006 Paris, France, [b]Institut de Mathématiques, University of Toulouse *Corresponding author. thierry.klein@math.univ-toulouse.fr

In this article, we consider the problem of defining a generalized notion of barycenter of random probability measures on $\mathbb{R}$. It is a well known fact that the set of Radon probability measures endowed with the 2-Wasserstein distance is not an Euclidean space. Consequently, to define a notion of barycenter for random probability measures, it is natural to use the notion of Fréchet mean [**?**] that is an extension of the usual Euclidean barycenter to non-linear spaces endowed with non-Euclidean metrics. If $\mathbb{Y}$ denotes a random variable with distribution $\mathbb{P}$ taking its value in a metric space $(\mathcal{M}, d_{\mathcal{M}})$, then a Fréchet mean (not necessarily unique) of the distribution $\mathbb{P}$ is a point $m^* \in \mathcal{M}$ that is a global minimum (if any) of the functional

$$J(m) = \frac{1}{2}\int_{\mathcal{M}} d^2_{\mathcal{M}}(m, y) d\mathbb{P}(y) \quad \text{i.e.} \quad m^* \in \arg\min_{m\in\mathcal{M}} J(m).$$

In this paper, a Fréchet mean of a random variable $\mathbb{Y}$ with distribution $\mathbb{P}$ will be also called a barycenter. For random variables belonging to nonlinear metric spaces, a well-known example is the computation of the mean of a set of planar shapes in the Kendall's shape space [**?**] that leads to the Procrustean means studied in [**?**]. Many properties of the Fréchet mean in finite dimensional Riemannian manifolds (such as consistency and uniqueness) have been investigated in [**?**, **?**, **?**, **?**, **?**, **?**].

This article is an attempt to use these tools and some extensions for analyzing computer codes outputs in a random environment, what is the subject of computer code experiments. In the first section we define new contrasts for random $c.d.f.$ by considering generalized " Wasserstein" costs. From this, in the second section we define new features in the way of the Fréchet mean that we call Fréchet features. Then we propose some examples. The next two sections are devoted to a sensitivity analysis of random $c.d.f.$, first from a Sobol point of view that we generalized to a contrast point of view as in [**?**].

1 Wasserstein distances and Wasserstein costs for unidimensional distributions

For any $p \geq 1$ we may define a Wasserstein distance between two distribution of probability, denoted F and G (their cumulative distribution functions, $c.d.f.$) on $\mathbb{R}^d$ by:

$$W_p^p(F, G) = \min_{(X,Y)} \mathbb{E}\|X - Y\|^p,$$

where the random variables ($r.v.$'s) have $c.d.f.$ F and G ($X \sim F, Y \sim G$), assuming that X and Y have finite moments of order p. We call Wasserstein$_p$ space the space of all $c.d.f.$ of $r.v.$'s with finite moments of order p.

As previously mentioned, in the unidimensional case where $d = 1$, it is well

known that $W_p(F,G)$ is explicitly computed by:

$$W_p^p(F,G) = \int_0^1 |F^-(u) - G^-(u)|^p du = \mathbb{E}|F^-(U) - G^-(U)|^p.$$

Here F^- and G^- are the generalized inverses of F and G that are increasing with limits 0 and 1, and U is a *r.v.* uniform on $[0,1]$. Of course $F^-(U)$ and $G^-(U)$ have *c.d.f.* F and G.

This result extends to more general contrast functions.

Definition 1.1 *We call contrast functions any application c from* $\mathbb{R}^2$ *to* $\mathbb{R}$ *satisfying the "measure property"* $\mathcal{P}$ *defined by*

$$\mathcal{P} : \forall x \leq x' \text{ and } \forall y \leq y', c(x',y') - c(x',y) - c(x,y') + c(x,y) \leq 0,$$

meaning that c defines a negative measure on $\mathbb{R}^2$.

Example 1.1 $c(x,y) = -xy$ *satisfies the* $\mathcal{P}$ *property.*

Remark 1 *If c satisfies* $\mathcal{P}$ *then any function of the form* $a(x) + b(y) + c(x,y)$ *satisfies* $\mathcal{P}$. *For instance* $(x-y)^2 = x^2 + y^2 - 2xy$ *satisfies* $\mathcal{P}$.

Remark 2 *More generally if C is a convex real function then c(x,y)=C(x-y) satisfies* $\mathcal{P}$. *This is the case of* $|x-y|^p$, $p \geq 1$.

Definition 1.2 *We define de Skorohod space* $\mathcal{D} := \mathcal{D}([0,1])$ *of all distribution functions that is the space of all non decreasing function from* $\mathbb{R}$ *to* $[0,1]$ *that are càd-làg with limit* 0 *(resp.* 1*) in* $-\infty$ *(resp.* $+\infty$*) equiped with the supremum norm.*

Definition 1.3 (The $c-$Wasserstein cost) *For any* $F \in \mathcal{D}$, *any* $G \in \mathcal{D}$ *and any positive contrast function c, we define the c− Wasserstein cost by*

$$W_c(F,G) = \min_{(X\sim F, Y\sim G)} \mathbb{E}\left(c(X,Y)\right) < +\infty$$

The following theorem can be found in ([**?**]).

Theorem 1.2 (Cambanis, Simon, Stout [?]) *Let c a function from* $\mathbb{R}^2$ *taking values in* $\mathbb{R}$. *Assume that it satisfies the "measure property"* $\mathcal{P}$. *Then*

$$W_c(F,G) = \int_0^1 c(F^-(u), G^-(u))du = \mathbb{E}\ c(F^-(U), G^-(U)),$$

where U is a random variable uniformly distributed on $[0,1]$.

At this point we may notice that in a statistical framework one encounter many contrasts that are defined via a convex function. Actually many features of probability distribution can be characterized via such a contrast function. For instance an interesting case is the quantiles. Applying the previous remark we get:

Proposition 1.1 *For any $\alpha \in (0,1)$ the contrast function (pinball function) associated to the α-quantile $c_\alpha(x,y) = (1-\alpha)(y-x)\mathbf{1}_{x-y<0} + \alpha(x-y)\mathbf{1}_{x-y\geq 0}$ satisfies $\mathcal{P}$.*

This result is the starting point of the definition of some new features of random *c.d.f.*.

2 Extension of the Fréchet mean to other features

A Fréchet mean $\mathcal{E}X$ of a *r.v.* X taking values in a metric space $(\mathcal{M}, d)$ is define as (whenever it exists):

$$\mathcal{E}X \in \operatorname{argmin}_{\theta\in\mathcal{M}} \mathbb{E}\ d(X,\theta)^2.$$

That means that it minimizes the contrast $\mathbb{E}\ d(X,\theta)^2$ which is an extension of the classical contrast $\mathbb{E}\|X-\theta\|^2$ in $\mathbb{R}^d$.

Adopting this point of view we can define a "Fréchet feature" associated to a convenient contrast function.

Now we consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a measurable application $\mathbb{F}$ from Ω to $\mathcal{D}$. Take c a positive contrast (satisfying property $\mathcal{P}$) and define the analogously to the Fréchet mean, the Fréchet feature associated to c or contrasted by c as it follows:

Definition 2.1 *Assume that $\mathbb{F}$ is a random variable taking values in $\mathcal{D}$. Let c be a non negative contrast function satisfying the property $\mathcal{P}$. We define a c-contrasted feature $\mathcal{E}_c\mathbb{F}$ of $\mathbb{F}$ by:*

$$\mathcal{E}_c\mathbb{F} \in argmin_{G\in\mathcal{D}} \mathbb{E}\left(W_c(\mathbb{F}, G)\right).$$

Of course this definition coincides with the Fréchet mean in the Wasserstein$_2$ space when using the "contrast function" $c(F,G) = W_2^2(F,G)$.

Theorem 2.1 *If c is a positive cost function satisfying the property $\mathcal{P}$, if the application defined on $(\omega,u) \in \Omega\times(0,1)$ by $\mathbb{F}^-(\omega,u)$ is measurable and if $\mathcal{E}_c\mathbb{F}$ exists and is unique we have:*

$$(\mathcal{E}_c\mathbb{F})^-(u) = argmin_{s\in\mathbb{R}} \mathbb{E}c(\mathbb{F}^-(u), s).$$

That is $\mathcal{E}_c\mathbb{F}$ is the inverse of the function taking value at u the c-contrasted feature of the real r.v. $\mathbb{F}^-(u)$. For instance the Fréchet mean in the Wasserstein$_2$ space is the inverse of the function $u \longrightarrow \mathbb{E}\ (\mathbb{F}^-(u))$.

Remark 3 *Here, we proposed a general framework on* $\mathbb{F}$ *and made some strong assumptions on existence uniqueness and measurability. But one can construct explicit parametric models for* $\mathbb{F}$*. We refer to [?] for such example. In particular in [?], the authors used some results of [?] that ensures measurability for some parametric models on* $\mathbb{F}$*.*

Another example is the Fréchet median. A contrast function defining the median in $\mathbb{R}$ is $|x-y|$. An immediate extension to the Wassertein$_1$ space is to consider the "contrast function" $c(F,G) = W_1(F,G)$. Thus we obtain the Fréchet median of a random $c.d.f.$ as :

$$(\mathrm{Med}(\mathbb{F}))^-(u) \in \mathrm{Med}(\mathbb{F}^-(u)).$$

More generally we can define an α-quantile of a random $c.d.f.$, $q_\alpha(\mathbb{F})$, as:

$$(q_\alpha(\mathbb{F}))^-(u) \in q_\alpha(\mathbb{F}^-(u)),$$

where $q_\alpha(X)$ is the set of the α-quantiles of X taking its values in $\mathbb{R}$.

Proof of Theorem 2.1.

Since c satisfies $\mathcal{P}$ we have:

$$\mathbb{E}\ W_c(\mathbb{F},G) = \mathbb{E}\int_0^1 c(\mathbb{F}^-(u),G^-(u))du = \int_0^1 \mathbb{E}\ c(\mathbb{F}^-(u),G^-(u))du,$$

by Fubini's theorem.

Now for all $u \in (0,1)$ the quantity $\mathbb{E}\ c(\mathbb{F}^-(u),G^-(u))$ is minimum for $G^-(u)$ a feature contrasted by c. Noticing that this results in an increasing and càd-làg function the theorem follows. □

3 Example

In this section we illustrate our definitions through an example.

Let F_0 an increasing absolutely continuous $c.d.f$ (hence F_0^{-1} exists), X a $r.v.$ with distribution F_0, M and Σ two real $r.v.$'s, $\Sigma>0$. We consider the random $c.d.f.$ $\mathbb{F}$ of $\Sigma X + M$. We have:

$$\mathbb{F}(x) = F_0(\frac{x-M}{\Sigma}) \text{ and } \mathbb{F}^{-1}(u) = \Sigma F_0^{-1}(u) + M.$$

As well known the Fréchet mean of $\mathbb{F}$ is given by: $(\mathcal{E}(\mathbb{F}))^{-1}(u) = \Sigma F_0^{-1}(u) + M$, thus $\mathcal{E}(\mathbb{F})(x) = F_0(\frac{x-\mathbb{E}M}{\mathbb{E}\Sigma})$.

Now using the α-quantile contrast $c_\alpha(x,y) = (1-\alpha)(y-x)\mathbf{1}_{x-y<0} + \alpha(x-y)\mathbf{1}_{x-y\geq 0}$ and following our definition, we define the "α-quantile" of $\mathbb{F}$:

$$(q_\alpha(\mathbb{F}))^{-1}(u) = q_\alpha(\Sigma F_0^{-1}(u) + M).$$

Assuming that $\Sigma = 1$ it simplifies in $q_\alpha(\mathbb{F})(x) = F_0(x - q_\alpha(M))$. When $M = 0$ we have $q_\alpha(\mathbb{F})(x) = F_0(\frac{x}{q_\alpha(\Sigma)})$ (see figure(**??**)).

Once these features defined, referring to computer experiment framework, in the next section we propose a sensitivity analysis of these Fréchet features of a random *c.d.f.* as stochastic output of a computer code.

4 Sensitivity indices for a random *c.d.f.*

4.1 Sobol index

A very classical problem in the study of computer code experiments (see [**?**]) is the evaluation of the relative influence of the input variables on some numerical result obtained by a computer code. This study is usually called sensitivity analysis in this paradigm and has been widely assessed (see for example [**?**], [**?**], [**?**] and references therein). More precisely, the numerical result of interest Y is seen as a function of the vector of the distributed input $(X_i)_{i=1,\cdots,d}$ $(d \in \mathbb{N}_*)$. Statistically speaking, we are dealing here with the unnoisy non parametric model

$$Y = f(X_1, \ldots, X_d), \tag{1}$$

where f is a regular unknown numerical function on the state space $E_1 \times E_2 \times \ldots \times E_d$ on which the distributed variables $(X_1, \ldots, X_d)$ are living. Generally, the inputs are assumed to be stochastically independent and sensitivity analysis is performed by using the so-called Hoeffding decomposition (see [**?**] and [**?**]). In this functional decomposition f is expanded as a L^2 sum of uncorrelated functions involving only a part of the random inputs. For any subset v of $I_d = \{1, \ldots, d\}$ this leads to an index called the Sobol index ([**?**]) that measures the amount of *randomness* of Y carried in the subset of input variables $(X_i)_{i\in v}$. Without loss of generality, let us consider the case where v reduces to a singleton. Let us first recall some well known facts about Sobol index. The global Sobol index quantifies the influence of the *r.v.* X_i on the output Y. This index is based on the variance (see [**?**],[**?**]): more precisely, it compares the total variance of Y to the expected variance of the variable Y conditioned by X_i,

$$S_i = \frac{\mathrm{Var}(\mathbb{E}[Y|X_i])}{\mathrm{Var}(Y)}. \tag{2}$$

By the property of the conditional expectation it writes also

$$S_i = \frac{\mathrm{Var}(Y) - \mathbb{E}(\mathrm{Var}[Y|X_i])}{\mathrm{Var}(Y)}. \tag{3}$$

In view of this formula we can define a Sobol index for the Fréchet mean of a random *c.d.f.* $\mathbb{F} = h(X_1, \ldots, X_d)$. Actually we define $\mathrm{Var}(\mathbb{F}) = \mathbb{E}W_2^2(\mathbb{F}, \mathcal{E}(\mathbb{F}))$, and

$$S_i(F) = \frac{\mathrm{Var}(\mathbb{F}) - \mathbb{E}(\mathrm{Var}[\mathbb{F}|X_i])}{\mathrm{Var}\mathbb{F}}.$$

From Theorem 2.1 we get:

$$\mathrm{Var}(\mathbb{F}) = \mathbb{E}\int_0^1 |\mathbb{F}^-(u) - \mathcal{E}(\mathbb{F})^-(u)|^2 du = \mathbb{E}\int_0^1 |\mathbb{F}^-(u) - \mathbb{E}\mathbb{F}^-(u)|^2 du = \int_0^1 \mathrm{Var}(\mathbb{F}^-(u)) du.$$

And the Sobol index is now:

$$S_i(\mathbb{F}) = \frac{\int_0^1 \mathrm{Var}(\mathbb{F}^-(u))du - \int_0^1 \mathbb{E}\mathrm{Var}[\mathbb{F}^-(u)|X_i]du}{\int_0^1 \mathrm{Var}(\mathbb{F}^-(u))du} = \frac{\int_0^1 \mathrm{Var}(\mathbb{E}[\mathbb{F}^-(u)|X_i])du}{\int_0^1 \mathrm{Var}(\mathbb{F}^-(u))du}.$$

As a toy example, applying this to our previous example $\mathbb{F}(x) = F_0(\frac{x-M}{\Sigma})$, where M and Σ play the role of influent random variables, we find:

$$S_\Sigma = \frac{\mathrm{Var}\ \Sigma + 2\mathrm{cov}(\Sigma, M)\mathbb{E}\xi}{\mathrm{Var}\ \Sigma + \mathrm{Var}\ M + 2\mathrm{cov}(\Sigma, M)\mathbb{E}\xi}, S_M = \frac{\mathrm{Var}\ M + 2\mathrm{cov}(\Sigma, M)\mathbb{E}\xi}{\mathrm{Var}\ \Sigma + \mathrm{Var}\ M + 2\mathrm{cov}(\Sigma, M)\mathbb{E}\xi}$$

where ξ has $c.d.f.$ F_0, since $\mathbb{E}\xi = \int_0^1 F_0^{-1}(u)du$.

In practice M and Σ depends upon numerous random variables $(X_1, \ldots, X_d)$, then the Sobol index with respect to X_i becomes:

$$S_i = \frac{\mathrm{Var}\ \mathbb{E}[\Sigma|X_i] + 2\mathrm{cov}(\mathbb{E}[\Sigma|X_i], \mathbb{E}[M|X_i])\mathbb{E}\xi + \mathrm{Var}\ \mathbb{E}[M|X_i]}{\mathrm{Var}\ \Sigma + \mathrm{Var}\ M + 2\mathrm{cov}(\Sigma, M)\mathbb{E}\xi}$$

4.2 Sensitivity index associated to a contrast function

The formula (3) can be extended to more general contrast functions. The contrast function naturally associated to the mean of a real $r.v.$ is $c(x, y) = |x-y|^2$. We have $\mathbb{E}Y = \mathrm{argmin}_{\theta\in\mathbb{R}}\mathbb{E}c(Y, \theta)$ and $\mathrm{Var}(Y) = \min_{\theta\in\mathbb{R}}\mathbb{E}c(Y, \theta)$. Thus the denominator of S_i is the variation between the minimum value of the contrast and the expectation of the minimum of the same contrast when conditioning by the $r.v.$ X_i. Hence for a feature of a real $r.v.$ associated to a contrast function c we defined a sensitivity index (see ([**?**])):

$$S_{i,c} = \frac{\min_{\theta\in\mathbb{R}}\mathbb{E}c(Y, \theta) - \mathbb{E}\min_{\theta\in\mathbb{R}}\mathbb{E}[c(Y, \theta)|X_i]}{\min_{\theta\in\mathbb{R}}\mathbb{E}c(Y, \theta)}.$$

Along the same line, we now define a sensitivity index for a c-contrasted feature of a random $c.d.f.$ by:

$$S_{i,c} = \frac{\min_{G\in\mathbb{W}}\mathbb{E}W_c(\mathbb{F}, G) - \mathbb{E}\min_{G\in\mathbb{W}}\mathbb{E}[W_c(\mathbb{F}, G)|X_i]}{\min_{G\in\mathbb{W}}\mathbb{E}W_c(\mathbb{F}, G)}.$$

The computation of $S_{i,c}$ simplifies when c satisfies the property $\mathcal{P}$ and assuming the uniqueness of $\mathcal{E}_c\mathbb{F}$:

$$S_{i,c} = \frac{\mathbb{E}\int_0^1 c(\mathbb{F}^-(u), (\mathcal{E}_c\mathbb{F})^-(u))du - \mathbb{E}[\int_0^1 c(\mathbb{F}^-(u), (\mathcal{E}_c[\mathbb{F}|X_i])^-(u))du]}{\mathbb{E}\int_0^1 c(\mathbb{F}^-(u), (\mathcal{E}_c\mathbb{F})^-(u))}$$

where $\mathcal{E}_c[\mathbb{F}|X_i]$ is the c-contrasted feature conditional to X_i (*i.e.* with respect to the conditional distribution of $\mathbb{F}$), also assumed to be unique.

For instance if $c = |x-y|$, $(\mathcal{E}_c\mathbb{F})^-(u)$ is the "median" (assumed to be unique) of the random variable $\mathbb{F}^-(u)$ and:

$$S_{i,\text{Med}} = \frac{\mathbb{E}\int_0^1 |\mathbb{F}^-(u) - Med(\mathbb{F}^-(u))|du - \mathbb{E}[\int_0^1 |\mathbb{F}^-(u) - Med[\mathbb{F}^-(u)|X_i]|du]}{\mathbb{E}\int_0^1 |\mathbb{F}^-(u) - Med(\mathbb{F}^-(u))|du}.$$

The same holds for any α-quantile, using the corresponding contrast function c_α but whith less readable formula.

5 Conclusion

This article is an attempt to define interesting features for a functional output of a computer experiment, namely a random $c.d.f.$, together with its sensitivity analysis. This theory is based on contrast functions that allow to compute Wasserstein costs. In the same way as the Fréchet mean for the Wassersstein$_2$ distance we have defined features that minimize some contrasts made of these Wasserstein costs. Straightforwardly from the construction of that features we have developed a proposition of sensitivity analysis, first of Sobol type and then extended to sensitivity indices associated to our new contrasts. We intend to apply our methodology to an industrial problem: the PoD (Probability of Detection of a defect) in a random environment. In particular we hope that our α-quantiles will provide a relevantt tool to analyze that type of data.